DOCUMENT RESUME

ED 202 024 CS 206 295

AUTHOR Crowhurst, Marion
TITLE The Effect of Syntactic Complexity on Writing Quality: A Review of Research.
PUB DATE [80]
NOTE 37p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Classroom Techniques; College Freshmen; *Difficulty Level; Elementary Secondary Education; Higher Education; Literature Reviews; *Sentence Combining; *Sentence Structure; *Syntax; Teaching Methods; Writing (Composition); Writing Exercises; Writing Instruction; *Writing Research; *Writing Skills

ABSTRACT

Research on the relationship between syntactic complexity and quality of written composition was reviewed to test the assumption that these two concepts are positively related. The studies reviewed, which have appeared during the past 15 years, fall into two broad categories: (1) intervention studies designed to increase syntactic complexity and to examine the effect of such increases on the quality of written composition, and (2) a set of diverse studies on the proposed relationship between syntactic complexity and writing quality. Based on the literature review, two conclusions were offered concerning the complexity-quality relationship. First, neither T-unit length nor clause length is a good predictor of writing quality. Second, although sentence-combining studies sometimes seem to improve writing quality, the improvement is probably due to factors other than increases in T-unit and clause length. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED202024

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

THE EFFECT OF SYNTACTIC COMPLEXITY ON WRITING QUALITY

A REVIEW OF RESEARCH

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Marion Crowhurst

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Author: Marion Crowhurst

Department of Language Education
University of British Columbia
2125 Main Mall
Vancouver, B.C. V6T 1Z5

CS206295

THE EFFECT OF SYNTACTIC COMPLEXITY ON WRITING QUALITY

A REVIEW OF RESEARCH

Various strands of research during the past fifteen years have seemed to indicate that syntactic complexity (i.e., such features as sentence length and clause length) is positively related to the quality of written composition. The purpose of this review is to examine that research and to draw such conclusions as may be warranted.

In a series of important studies, Kellogg Hunt (1965, 1966, 1970) demonstrated that syntactic complexity in written composition increases with age. He used various measures of syntactic complexity--sentence length, clause length, ratio of subordinate clauses to all clauses, and a new measure, the minimum terminable unit or T-unit which he defined as "one main clause plus any subordinate clause or nonclausal structure that is attached to or embedded in it"[1]. (Hunt, 1970, p.4). In mature writing a T-unit is usually the same as a sentence. However, immature writing is frequently marked by excessive coordination of main clauses so that sentences such as the following occur: "The boys put on their jackets and they went out of the house and they began to play hockey on the road but their father told them it was dangerous." The sentence just cited contains four T-units, i.e., four main clauses. The length of the sentence is thirty words; mean T-unit length is 7.5 words.

Hunt concluded that mean T-unit length and mean clause length were the best measures of "syntactic maturity." This conclusion was supported by O'Donnell, Griffin and Norris (1967), O'Donnell (1976) and Loban (1976) who found high positive correlations between T-unit length (Loban's communication unit) and more refined and discriminating measures of syntactic development which were much more expensive and difficult to use and which appeared to offer little advantage in measuring the syntactic development of large groups of children. T-unit length and clause length have come to be widely used as measures of syntactic development not only throughout the English-speaking world, but in other languages as well (Reesink, Holleman-van der Sleen, Stevens, & Kohnstumm, 1971; Wijnstra, 1972).

Hunt recommended that curriculum materials be designed to facilitate the ability to combine short, simple sentences into longer sentences in the hope that the syntactic development which occurred naturally with age could be facilitated by instruction. One result of his research and the recommendations based thereon has been the development of sentence-combining curricula. Numerous studies of sentence-combining in the past decade (e.g., Combs, 1976; Miller and Ney, 1968; Mellon, 1969; Morenberg, Daiker & Kerek, 1978; O'Hare, 1973) have shown beyond question that instruction in sentence-combining produces increases in syntactic complexity in written composition over a wide age-range (Grade 4 to college).

Hunt (1966,1970) found not only that syntactic complexity scores increased with age, but also that students of high ability had significantly higher scores on clause and T-unit length than those of average and low ability, a finding supported by the results of several subsequent studies (Loban, 1976; Richardson, Calnan, Essen and Lambert, 1976; San Jose, 1972). It was a short step from this finding to the hypotheses that clause and T-unit length might be positively associated with quality of written composition and that sentence-combining, which increases syntactic complexity, might also improve written composition. These hypotheses have been explored in a number of studies. The purpose of this article is to review those studies and to determine what conclusions may be drawn from their findings. Two kinds of studies will be examined: a. a set of diverse studies which have examined the relationship between syntactic complexity and quality (prediction/relationship studies); and, b. intervention studies designed to increase syntactic complexity through instruction and to examine the effect of such instruction on the quality of written composition.

Prediction/Relationship Studies

In a small exploratory study, Potter (1967) compared the twenty best and the twenty worst out of one hundred tenth-grade expository compositions and reported that mean T-unit length was greater in the good compositions (16.00) than in the poor (14.2). No statistical analysis was done.

Out of eighty-eight freshman compositions describing an important person, Schmeling (1969) identified fifteen "good" compositions, fifteen "average" compositions, and fifteen "poor" compositions on the basis of holistic ratings by three raters. He found that good compositions had a mean T-unit length (15.04) and a mean clause length (8.47) slightly greater than the poor compositions (Wds/TU = 14.47; Wds/CL = 8.13), but that the average compositions had higher scores than either good or poor (Wds/TU = 16.34; Wds/CL - 9.05). An analysis of variance failed to find these differences significant. However, the small number of subjects may have reduced the power of the test to detect small differences.

Rosen (1969) examined the writing of fifty British students aged 15-16 years. Eight different sets of topics were formulated, each set designed to elicit a different kind of writing. Each student wrote one composition on a topic he/she selected from each of the eight sets. Compositions were given holistic quality ratings by nine raters. Rosen found that the composition set which had the lowest average T-unit length for his fifty subjects, (i.e., Set 6 which elicited narrative writing) received the highest mean quality score and the set which had the highest average T-unit length (Set 2 which elicited expository writing) had the lowest mean quality score of all eight sets. When mean T-unit length was averaged over all eight composition sets, his best writers (Group 1) had the highest mean T-unit length (15.58) and his worst writers had

the lowest mean T-unit length (14.01), but this was not consistently true for each of the eight topic sets individually. On Set 6 (narrative), for example, his best writers had the second lowest mean T-unit score of all five groups. Rosen's work is descriptive with no statistical analysis and is chiefly interesting in the questions it raises and the indication it gives of a complicated relationship between T-unit length and quality.

Gebhard (1978) selected thirty-three "poor" and twenty-one "good" essays from 500 expository essays written by college freshmen. Using a variety of measures including T-unit length and clause length, she performed T-test analyses to compare good and poor freshman essays with each other and each of these groups with expository prose written by twenty-five professional writers. Both T-unit length and clause length were significantly longer for the professional essays than for good freshman ($p < .001$) and for poor freshman essays ($p < .001$). Clause length for good freshman essays was significantly longer ($p < .05$) than for poor freshman essays, but there was no significant difference between good and poor freshmen on T-unit length. Gebhard concludes that clause length appears to be a better indicator of quality in freshman essays than T-unit length.

Two further findings of Gebhard's are worth noting. Mean number of words per sentence combining transformation distinguished

between good and poor freshmen and between professional writers and each group of freshman writers ($p < .001$). Secondly, professional essays were distinguished from freshman essays not only by clause and T-unit length, but also by high standard deviations on both these measures (Wds/CL: $\bar{x} = 14.30$, sd = 14.72; Wds/TU: $\bar{x} = 20.75$, sd = 11.9). Gebhard gives no standard deviations for her freshmen, but both Faigley (1979) and Morenberg et al. (1978) report standard deviations for freshmen of less than three words on T-unit length for means of 14 - 15 words and standard deviations of less than 1.5 words on clause length for means of 7.7 - 8.8 words. The high standard deviations of Gebhard's professional writers indicate the facility with which the professional writer lengthens or shortens his utterance.

Crowhurst (1980b) examined the relationship between T-unit length and quality ratings of narrative and argumentative compositions of pupils in Grades 6, 10 and 12. Having concluded in earlier studies (Crowhurst and Piche, 1979; Crowhurst, 1980a) that argument places greater demands on writers to use their syntactic resources than does narration, she predicted that agruments of high syntactic complexity would receive higher quality ratings than arguments of low syntactic complexity by the same students, but that this would not be the case for narrations. More than 200 subjects at each grade level were randomly assigned to write either three narratives or three

arguments. From these triads, she selected pairs of compositions, one of high syntactic complexity and one of low syntactic complexity, each pair written by the same student. Pairs of compositions were selected if: a. they were of comparable length; b. they differed by a minimum of two words on mean T-unit length; and, c. one had a mean T-unit length at least .5 higher and the other a mean T-unit length at least .5 lower than the mean for all compositions in the same mode and grade. The number of pairs thus obtained ranged from sixteen in Grade 10 narration to twenty-nine in Grade 12 argument. Compositions at each grade level were quality-rated by four experienced raters. Quality scores were analyzed by analysis of variance. As predicted, arguments of high syntactic complexity received significantly higher quality scores than arguments of low syntactic complexity at both Grades 10 and 12. There was no significant difference at Grade 6. Crowhurst suggests that the difficulty Grade 6 students experience in writing argumentation (Crowhurst, 1978) may have been a complicating factor in evaluating Grade 6 arguments. Also as predicted, narratives of high syntactic complexity did not receive significantly higher quality scores than narratives of low syntactic complexity. Indeed, at Grade 12, narratives of <u>low</u> syntactic complexity scored significantly higher than narratives of high syntactic complexity.

Three studies used multiple regression techniques to examine the relationship between various measures of syntactic complexity and quality scores (Faigley, 1979; Nold and Freedman, 1977; and Stewart and Grobe, 1979). Nold et al. collected four expository essays from each of twenty-two college freshmen and had six experienced raters assign a holistic quality score to each. They performed a stepwise multiple regression using quality as the criterion variable and seventeen predictor variables including words per T-unit, words per main clause, and words per subordinate clause. Neither T-unit length nor clause length was a significant predictor of quality. The only predictor variable positively associated with quality at a significant level was the precentage of words in final modifiers.

Faigley (1979) collected two narrative compositions from each of 138 college freshmen. He performed a stepwise multiple regression using a holistic quality rating as the criterion variable and six predictor variables: the percentage of T-units with final free modifiers, composition length, words per T-unit, words per clause, clauses per T-unit, and the percentage of words in final free modifiers. The six predictor variables explained only 22 percent of the variance of the quality ratings. The most influential of the six was the percentage of T-units with final free modifiers which explained

sixteen percent of the variance. Words per T-unit and words per clause together explained only one percent of the variance in quality ratings.

Stewart and Grobe (1979) examined the relationship between quality and various factors of syntactic complexity and mechanics of writing at Grades 5, 8 and 11. A ten percent subsample was randomly selected from a province-wide sample in New Brunswick, the subsample consisting of eighty-five Grade 5 students, eighty Grade 8 students and sixty-seven Grade 11 students. One expository composition by each subject was scored by two raters using a four-point holistic scale. A step-wise multiple regression was performed with the quality rating as the criterion variable and the following predictor variables: composition length, words per T-unit, words per clause, clauses per T-unit, and nine factors relatint to mechanics of writing. At Grade 5, words per T-unit and words per clause explained respectively eleven percent and three percent of the variance in quality scores. At Grades 8 and 11, words per T-unit and words per clause, together, explained only one percent of the quality variance.

Table 1 presents the simple correlations between quality scores and words per T-unit and words per clause, respectively, for each of the three regression studies described above. There

was a significant but weak correlation between words per T-unit and quality at Grade 5 (Stewart et al.) ($r = .30$; $p < .05$), but none at Grade 8, Grade 11 (Stewart et al.) or at the college freshman level (Faigley; Nold et al.). There were weak, significant correlations between words per clause and quality at Grade 5 (Stewart et al.) ($r = .23$, $p < .05$), and for Faigley's college freshmen ($r = .18$, $p < .05$), but not at Grade 8 or Grade 11 (Stewart et al.) or for Nold et al.'s college freshmen.

Insert Table 1 about here

Intervention Studies

Intervention studies follow a pre-test, intervention, post-test design in which the pre- and post-tests are written compositions and the intervention is a series of lessons and practice exercises designed to increase the syntactic complexity of subjects' sentences. In most studies, the intervention procedure involves teaching students to combine given simple sentences into longer sentences using specified syntactic constructions. For example, a problem which reads thus:

i. SOMETHING startled Tony.

ii. The shutter banged. (ING + OF)

instructs the student to combine the given sentences by transforming the verb in sentence ii to the -ing form so as to produce the sentence, The banging of the shutter startled Tony.

Two studies (Hake and Williams, 1979; Faigley, 1979) used different methods. Hake and Williams' sentence imitation procedure required students to produce a sentence following the structure of a given model sentence but providing their own content. Faigley used Francis Christensen's generative rhetoric in which students are given a base clause and instructions to add modifiers of specified kinds. Students might be given the base clause, a child entered the church, and told to add: a. two pre-nominal modifiers describing the child's appearance, and b. two participial phrases each describing something the child was doing. These instructions might produce such sentences as: A small, wide-eyed child entered the church singing lustily, clasping a large red nymnal in her hand, or, A scrawny, bare-foot child entered the church, walking carefully on the rich, red carpet, feeling the thick pile first with one dirty foot and then the other.

Numerous studies have shown beyond reasonable doubt that intervention studies such as those described above can significantly increase the syntactic complexity of written composition over a wide age-range from elementary school through college freshmen. The question examined in the present review is whether such interventions lead also to an improvement in the quality of that writing.

Mellon (1969) conducted a sentence-combining intervention study involving 247 seventh-grade students in twelve different classes taught by eleven different teachers at four different schools. Five experimental classes studied transformational grammar and sentence-combining, five control classes studied traditional grammar and usage from the regular textbook, and two placebo groups studied no grammar at all, but spent extra time on composition. All classes did required composition material from the common assigned text. To examine quality, Mellon randomly selected thirty-five students from each of experimental, control and placebo groups in a single school, only one school having allowed a placebo group. Two pre-test and two post-test descriptive compositions were each rated by three raters using a five-point holistic scale. The post-text quality scores of the three groups were compared by analysis of variance with pre-test scores covaried. Quality scores for the control group were significantly higher than for either the experimental or the placebo group. Mellon attributed the better performance of the control group to the teacher variable, noting that both control classes were taught by a single teacher who was the most highly experienced teacher involved in the study (p. 70).

O'Hare's (1973) sentence-combining intervention study involved eighty-three seventh-grade subjects in a single Florida high school who were randomly assigned to two control and two experimental classes. O'Hare and another experienced teacher each taught one experimental and one control class. Both groups followed the same English curriculum except that experimental classes spent less time on the regular program so that one and a half hours a week from mid-October to mid-May could be spent on sentence-combining instruction and practice. To compare composition quality of experimental and control subjects, O'Hare used a system of forced choices between compositions written by matched pairs of subjects. Thirty control subjects were randomly chosen and each was matched with an experimental subject of the same sex and approximately the same IQ (California Test of Mental Maturity). The thirty matched pairs were divided into two balanced groups of fifteen. Post-test quality scores on Composition 1 (a narration) were compared for one group of fifteen pairs, and scores on composition 2 (a description) were compared for the other group. Eight experienced evaluators selected the better of each of the thirty pairs of compositions, basing their decision on ideas, organization, style, vocabulary and sentence structure. Using a chi-square analysis of raters' judgments, O'Hare found that experimental compositions were selected significantly more frequently than control compositions ($p < .001$).

Combs (1976) conducted a sentence-combining intervention involving two experimental and two control seventh-grade classes in a single school (N = 100). Two teachers each taught one experimental and one control class, all classes following a single curricular format excepting for the sentence-combining exercises completed by the experimental classes. Combs used a matched pairs sampling strategy to compare writing quality. Twenty-two control subjects were matched for sex and ability (Lorge-Thorndike Ability Test) with twenty-two experimental subjects. Seven raters chose between matched pairs of pre-test and matched pairs of post-test compositions in narrative and descriptive modes. Each chosen composition received a score of one so that each composition received a score of from 0 to 7 depending on the number of raters who chose it. T-test comparisons between control and experimental groups showed no significant difference on quality scores at pre-test time. At post-test time, experimental compositions received significantly higher quality ratings than control compositions ($p < .05$). Combs (1977) conducted a delayed post-test eight weeks after the post-test and found that significantly higher scores had been maintained for experimental compositions.

Callaghan and Sullivan (Sullivan, 1979) conducted parallel, year-long sentence-combining studies at Grades 9 and 11 respectively involving nine Grade 9 teachers and

seven Grade 11 teachers teaching four classes each--one control and three experimental classes. Post-test expository compositions by twenty-five randomly selected experimental subjects were compared with compositions of twenty-five randomly selected control subjects at each grade level. At Grade 9, each essay was evaluated by four raters using an analytic scale on five criteria. At Grade 11, each essay was rated by four raters using a four-point holistic score. There was no significant difference between experimental and control quality scores at either grade level.

Morenberg, Daiker and Kerek (1978) conducted a fifteen-week (i.e., one term) sentence-combining study in which 290 college freshman composition students were randomly assigned to six experimental and six control sections. Teachers chose to teach either an experimental or a control class. The groups of experimental and control teachers were comparable in experience, qualifications, and assessed teaching effectiveness. Control sections studied rhetoric, analyzed professionally written essays and discussed student compositions. Experimental sections engaged in sentence-combining activities with considerable emphasis on discussion of the rhetorical reasons for choosing one pattern of sentence-combining over another. All pre-test and post-test expository compositions were rated for quality by four raters using a six-point holistic scale and by four different raters using an analytic

scale on six criteria. In addition, the forced-choice method was used on 134 experimental-control pairs of post-test compositions matched according to the writers' pre-test holistic scores. Post-test scores were tested by an analysis of covariance with the pre-test scores as covariates. At pre-test time, there was no significant difference between experimental and control quality scores. At post-test time, the experimental group scored significantly higher than the control group on both the holistic score ($p < .001$) and on five out of the six factors of the analytic scale (significance levels ranging from .05 to .001). In the forced-choice rating, experimental papers were chosen as better significantly more often than the control papers ($p < .05$).

Faigley (1979) conducted a one-term intervention in which instruction in Christensen's generative rhetoric was used to try to increase syntactic fluency and to improve the writing quality of college freshmen. One hundred and thirty-eight subjects were randomly assigned to four experimental sections and four control sections. Control sections followed a standard freshman composition course involving the study of rhetoric and discussion of essays written by skilled writers. Experimental sections worked on exercises on generative rhetoric. The four experimental and the four control teachers were carefully matched as to experience, qualifications and assessed teaching effectiveness. All pre-

and post-test narrative compositions were rated for quality by five raters using a six-point holistic scale. Scores were analyzed by means of a one-way analysis of covariance with pre-test scores as covariates. At pre-test time, there was no significant difference between experimental and control quality scores. At post-test time, experimental scores were significantly higher than control ($p < .01$).

Hake and Williams (1979) conducted an intervention involving 244 tenth-graders in which two different methods were used to try to increase syntactic fluency. Five classes of students in two different high schools engaged in approximately eighteen weeks of sentence-combining instruction; five classes in the same two high schools engaged in eighteen weeks of sentence-imitation instruction; two classes, one in each of two additional high schools, were used as a control group.

The experimenters categorized the pre-test expository essays as either "competent" or "incompetent" on the basis of flaws in organization, coherence, usage and punctuation. When they calculated words per T-unit and words per clause for each group, they found that the 145 incompetent students had significantly higher mean T-unit length ($p < .05$) than the ninety-nine competent students. They categorized the 212 post-test expository essays as competent or incompetent and

compared clause length and T-unit length for three groups of students: a. "competent" at pre-test who remained "competent" at post-test (N = 99); b. "incompetent" who remained "incompetent" (N = 37); and c. "incompetent" who became "competent" (N = 76). At post-test time, the following results pertained:

a. For both sentence-combiners and sentence-imitators, competent students who remained competent significantly increased mean T-unit length (from 11.86 to 15.56) and mean clause length (from 7.05 to 8.54).

b. For both sentence-combiners and sentence-imitators, incompetent students who became competent significantly decreased mean T-unit length (from 15.69 to 13.01).

c. Sentence-imitators who were incompetent and remained incompetent showed no significant difference in mean T-unit length or mean clause length.

d. Sentence-combiners who were incompetent and remained incompetent significantly increased mean T-unit length (from 14.92 to 17.11).

In this study, it appears that both sentence-combining ans sentence-imitation produced increases in T-unit and clause length for competent students and that the increases were not associated with any increases in composition flaws. Both sentence-combining and sentence-imitation produced decreases in T-unit length and clause length for a group of

incompetent students, and these decreases were associated with a decrease in the number of composition "flaws". There was a group of incompetent students whose already inflated T-unit counts were further increased by sentence-combining instruction and the increases were not associated with any improvement in the number of composition flaws noted.

The liklihood of grammatical and logical errors increases with increases in the amount of coordination and subordination within a sentence. It is to be expected, then, that some students may show higher error rates if they increase T-unit length. The danger of increased error rates is especially great if students who are already prone to make such errors deduce from sentence-combining instruction that long sentences (or T-units) are desired and valued by the teacher. It would appear that this may have occurred with some of Hake and Williams' originally incompetent students.

Obenchain (1979) conducted a sentence-combining study with sixty-one tenth graders in which the focus was on paragraphy unity, coherence and emphasis through the identification of supporting and non-supporting ideas and the proper use of connectives. She found, at post-test time, dramatic decreases in punctuation, grammatical and spelling errors.

DISCUSSION

The intervention studies described above allow of a cautious conclusion that instruction such as sentence-combining

which aims at--and succeeds in--increasing syntactic fluency may also lead to an improvement in over-all writing quality. The Morenberg et al. (1978) and Faigley (1979) studies provide convincing evidence that, at the freshman college level, the quality of written composition may be improved by composition programs which focus on increasing syntactic fluency. It is to be noted that both studies involved open, "whole discourse" problems rather than cued exercises and single-sentence problems. There was, moreover, considerable discussion of the rhetorical effect of the various suggested answers to the problems posed. The programs thus involved considerably more than merely increasing syntactic fluency.

At the Grade 7 level, O'Hare (1973) and Combs (1976) found experimental post-test compositions to be significantly higher on quality scores than control compositions, but Mellon's (1969) control group scored significantly higher on quality at post-test time than either the experimental or the placebo group. Mellon attributes this result to a teacher effect since the control group was taught, as he subsequently discovered, by a very experienced teacher who was reputed to be the best composition teacher in the school (Mellon, 1979, p. 15). (It is interesting to note that Mellon did not expect his sentence-combining group to improve in writing quality and consistently maintains that relatively short-term sentence-combining interventions should not be expected to

produce immediate, over-all quality improvement (1969, 1979). His quality check was designed to make sure that the writing of his experimental group was not adversely affected by sentence-combining instruction.)

It is possible that teacher behaviour may have been an influential factor in the positive results obtained by O'Hare and Combs. In both cases, only two teachers were involved, each teacher instructing both an experimental and a control group. O'Hare himself was one of the teachers in his study. Students' positive response to sentence-combining has been widely reported. Teachers engaging in a new and promising curricular activity for which they needed special instruction (Combs, private communication) and which was received positively by students, might be expected to approach those classes with a degree of enthusiasm which, in itself, might make for positive results. On the basis of a pilot study, Faigley (1979) decided against his original plan of having each teacher teach both a control and an experimental section. The decision was based partly on the difficulty of having teachers approach both sections with equal enthusiasm (p.199). Greater confidence might be placed in the O'Hare and Combs findings if a larger number of teachers had been involved. Callaghan and Sullivan (Sullivan, 1979), whose studies involved many teachers (nine and seven, respectively), each teaching three experimental classes and one control class, did not find

writing quality to be superior in their experimental group. The conclusion that instruction designed to increase syntactic fluency is likely to produce also an improvement in over-all writing quality should be taken with appropriate caution. It is to be noted that the studies which speak most convincingly (Faigley, 1979; Morenberg et al., 1978) involved considerably more than mere practice in syntactic forms.

Even if it be tentatively concluded that intervention activities such as those described lead to increases both in syntactic fluency and in over-all quality, the question of causality remains. Is the improvement in quality related to increases in syntactic complexity scores, and, in particular, to the two most commonly used measures of syntactic complexity--T-unit length and clause length?

T-unit and clause length have sometimes been found to be greater in "better" compositions than in "poorer" and in the writing of those of superior ability (Gebhard, 1978; Hunt, 1966; Loban, 1976; Potter, 1967; Richardson et al, 1976; San Jose, 1972). However, these findings do not necessarily imply a causal relationship especially in view of the fact that a positive relationship between these measures and quality has not always been found. While Rosen's (1969) best students had the greatest mean T-unit

length when summed over all kinds of writing, they did not have the highest T-unit length for every writing set (e.g., Set 6 which elicited narrative writing). Hake and Williams' (1978) originally incompetent students had greater T-unit and clause lengths than their originally competent students, and those who remained incompetent after sentence-combining instruction actually increased their already high scores for T-unit and clause length. T-unit length was not greater (though clause length was) for Gebhard's "good" college freshmen than for her "poor" college freshmen.

No significant correlation was found between quality and T-unit length by Faigley (1979) or by Nold et al. (1977) at the college level, or by Stewart et al. (1979) at Grades 8 and 11. Only at Grade 5 was there a significant positive correlation between T-unit length and quality (Stewart et al., 1979). A significant correlation between quality and clause length was found by Faigley at the college level and by Stewart et al. at the Grade 5 level, but no correlation was found by Nold et al. at the college level or by Grobe et al. at Grades 8 and 11.

Two conclusions seem warranted. One is that there is considerable variability in the relationship between T-unit length and quality. While it seems true that those of

superior ability usually have longer than usual T-unit scores, especially when summed over many pieces of writing of various kinds, it is also true that greater T-unit length is sometimes associated with faulty writing (Hake et al., 1979). Moreover, the relationship between T-unit length and quality appears to vary with the mode of writing. Crowhurst (1980) found that arguments of greater T-unit length were assigned higher quality ratings than those with shorter T-unit length, but that narratives of greater T-unit length were not assigned higher quality ratings, indeed, at the Grade 12 level, narratives of lower T-unit length were assigned higher quality scores than those of greater T-unit length. Similarly, Rosen found that for narrative writing, his good writers had one of the lowest mean T-unit scores of the five groups of students.

The second conclusion is that clause length appears to be only slightly better as a predictor of writing quality than T-unit length. It distinguished between Gebhard's (1978) good and poor college freshmen whereas T-unit length did not; it showed a small but significant positive correlation with writing quality for Faigley's (1979) college freshmen, whereas T-unit length did not. However, there is no dependable correlation between clause length and quality. Moreover, the amount of the variance in quality scores predicted by clause length and T-unit length was minute in

Faigley's study, in Nold et al.'s study (1977), and in Stewart et al.'s (1979) at Grades 8 and 11.

While T-unit length and clause length have proved to be useful, if somewhat gross, measures for charting the development of syntactic fluency in large numbers of subjects, they are not sufficiently discriminating to serve as reliable predictors of quality. These measures do not distinguish among the ways in which length is achieved. A measure is needed which will be sensitive to different levels of linguistic maturity in the structures within the T-unit or the clause. Measures which emerge as possibly more useful from the studies surveyed here are: the number of words per sentence-combining transformation which distinguished at the .001 level between Gebhard's (1978) good and poor college students, and between her professional writers and each group of freshman writers; and, secondly, measures related to final free modifiers. Faigley (1979) found that the percentage of T-units containing final free modifiers correlated .41 with quality and accounted for sixteen percent of the variance of quality scores; Nold et al. (1977) found that the percentage of words in final free modifiers correlated .42 with quality and explained twelve percent of the variance of quality scores. A problem with final free modifiers is that they occur infrequently in the writing of even college students. For

example, they occurred in only forty-one out of eighty-eight essays of Nold et al.'s college freshmen. Other more discriminating measures of syntactic fluency have been suggested by Belanger (1978), Botel and Granowsky (1972), Endicott (1973), Golub and Kidder (1974), and DiStefano and Howie (1979).

In view of the evidence that T-unit length and clause length are poor predictors of composition quality, it seems reasonable to suppose that the improvements in quality which sometimes accompany sentence-combining instruction may be due to factors other than increases in T-unit length and clause length. Other reasons for the success of sentence-combining and sentence-imitation programs have been suggested. Kinneavy (1979) points out that sentence-combining subjects are constantly writing sentences and suggests that such writing improvement as occurs may result from disciplined writing practice. Stotsky (1975) suggests that the practice of playing mentally and operationally with syntactic structures may lead "to a kind of automization of syntactic skills such that mental energy is freed in a Brunerian ...sense to concentrate on greater elaboration of intention and meaning" (p. 55). Mellon (1979) suggests that sentence-combining, when based on non-cued, whole-discourse exercises, "lowers writing anxiety by allowing the writer to deal with ... content in which he or she has no stake, and thus engenders respect

for and confidence in the student's own syntactic resources and strategies" (pp.28-29).

The conclusions emerging from the studies here reviewed have implications for teachers in the field. Teachers may use sentence-combining, sentence-expanding and sentence-imitation exercises with well-founded expectations that such instruction will lead to the development of students' syntactic resources. Over the long term, a sentence-combining strand in the composition program may contribute to an improvement in writing quality. However, increases in over-all writing quality should not be expected to result from short-term instructional programs since composing is a complex act which depends on many skills. Sentence-combining programs are most likely to contribute to over-all improvement in writing quality if students do substantial work on open or non-cued, whole-discourse problems, and if substantial attention is paid to discussing the rhetorical effect of the various versions which students produce. Teachers should be aware that sentence-combining instruction may lead to excessively long and awkward T-units in the writing of some children. The writing of superior, professional writers is marked not only by high average T-unit length but also by great variation in the length of their T-units. Teachers should therefore value and encourage not so much length of T-unit, but variety in both T-unit length and T-unit structure. Finally, teachers should not be misled into supposing that compositions are

better or more "mature" simply because mean T-unit length is great, especially if the compositions are narratives.

REFERENCES

Belanger, J.F. Calculating the syntactic density score: A mathematical problem. Research in the Teaching of English, 1978, 12, 149-53.

Botel, M., & Granowsky, A. A formula for measuring syntactic complexity: A directional effort. Elementary English, 1972, 49 (4), 513-516.

Combs, W. Further effects of sentence-combining practice on writing ability. Research in the Teaching of English, 1976, 10, 137-149.

Combs, W.E. Sentence-combining practice: Do gains in judgments of writing 'quality' persist? Journal of Educational Research, 1977, 70, 318-321.

Crowhurst, M. The effect of audience and mode of discourse on the syntactic complexity of the writing of sixth and tenth graders (Doctoral dissertation, University of Minnesota, 1977). Dissertation Abstracts International, 1978, 38, 7300A-7301A.

Crowhurst, M. Syntactic complexity in narration and argument at three grade levels. Canadian Journal of Education, 1980 (a), 5, 6-13.

Crowhurst, M. Syntactic complexity and teachers' quality ratings of narrations and arguments. Research in the Teaching of English. 1980 (b), 14, 223-231.

Crowhurst, M., & Piche, G.L. Audience and mode of discourse effects on syntactic complexity in writing at two grade levels. Research in the Teaching of English, 1979, 13, 101-110.

DiStefano, P., & Howie, S. Sentence weights: An alternative to the T-unit. English Education, 1979, 11, 98-102.

Endicott, A.L. A proposed scale for syntactic complexity. Research in the Teaching of English, 1973, 7, 5-12.

Faigley, L. The influence of generative rhetoric on the syntactic maturity and writing effectiveness of college freshmen. Research in the Teaching of English, 1979, 13, 197-206.

Gebhard, A.O. Writing quality and syntax: a transformational analysis of three prose samples. Research in the Teaching of English, 1978, 12, 211-232.

Golub, L.S., & Kidder, C. Syntactic density and the computer. Elementary English, 1974, 51, 1128-1131.

Hake, R., & Williams, J.M. Sentence expanding: not can, or how, but when. In D. Daiker, A. Kerek & M. Morenberg (Eds.), Sentence combining and the teaching of writing. Akron, Ohio: L & S Books, 1979.

Hunt, K.W. Grammatical structures written at three grade levels. Research Report No. 3, Champaign, Ill.: National Council of Teachers of English, 1965.

Hunt, K.W. Sentence structures used by superior students in grades four and twelve and by superior adults. Tallahassee: Florida State University, 1966.

Hunt, K.W. Syntactic maturity in school children and adults. Monographs of the Society for Research in Child Development, 1970, 35, (serial No. 134).

Kinneavy, J.L. Sentence combining in a comprehensive language framework. In D. Daiker, A.Kerek & M. Morenberg (Eds.), Sentence combining and the teaching of writing. Akron, Ohio: L & S Books, 1979.

Loban, W. Language development: kindergarten through grade twelve. Research Report No. 18, Urbana, Ill.: National Council of Teachers of English, 1976.

Mellon, J.C. Transformational sentence-combining. Research Report No. 10. Champaign, Ill.: National Council of Teachers of English, 1969.

Mellon, J.C. Issues in the theory and practice of sentence-combining: A twenty-year perspective. In D. Daiker, A. Kerek & M. Morenberg (Eds.) Sentence combining and the teaching of writing. Akron, Ohio: L & S Books, 1979.

Miller, B.D., & Ney, J.W. The effect of systematic oral exercises on the writing of fourth-grade students. Research in the Teaching of English, 1968, 2, 44-61.

Morenberg, M., Daiker, D., & Kerek, A. Sentence combining at the college level: an experimental study. Research in the Teaching of English, 1978, 12, 245-256.

Nold, E.W., & Freedman, S. An analysis of readers' responses to essays. Research in the Teaching of English, 1977, 11, 164-174.

Obenchain, Anne. Developing paragraph power through sentence combining. In D. Daiker, A. Kerek & M. Morenberg (Eds.) Sentence combining and the teaching of writing. Akron, Ohio: L & S Books, 1979.

O'Donnell, R.C. A critique of some indices of syntactic maturity. Research in the Teaching of English, 1976, 10, 31-38.

O'Hare, F. Sentence-combining: improving student writing without formal grammar instruction. Research Report No. 15. Urbana, Ill.: National Council of Teachers of English, 1973.

Perron, J.D. Written syntactic complexity and the modes of discourse. Paper presented at the annual meeting of the American Educational Research Association, New York, April, 1977.

Potter, R.R. Sentence structure and prose quality. Research in the Teaching of English, 1967, 1, 17-28.

Reesink, G.P., Holleman-van der Sleen, S.B., Stevens, K., & Kohnstumm, G.A. (Development of syntax among schoolchildren and adults: a replication-investigation.) Nederlands Tydschrift voor de Psychologie en haar Grensgebieden, 1971, 26, 335-364. (Psychological Abstracts, 1971, 47, No. 10536.)

Richardson, K., Calnan, M., Essen, J., & Lambert, L. The linguistic maturity of 11-year olds: Some analyses of the written compositions of children in the National Child Development Study. Journal of Child Language, 1976, 3, 99-115.

Rosen, H. An investigation of the effects of differentiated writing assignments on performance in English composition of a selected group of 15 and 16 year old pupils. Unpublished doctoral dissertation, University of London, 1969.

San Jose, C.P.M. Grammatical structures in four modes of writing at four grade levels (Doctoral dissertation, Syracuse University, 1972). Dissertation Abstracts International, 1973, 33, 5411A. (University Microfilms No. 73-9563).

Schmeling, H.H. A study of the relationship between certain syntactic features and overall quality of college freshman writing. (Doctoral dissertation, George Peabody College for Teachers, 1969). Dissertation Abstracts International, 1970, 30, 4970A. (University Microfilms No. 70-7641).

Stewart, M.F., & Grobe, C.H. Syntactic maturity and mechanics of writing: their relationship to teachers' quality ratings. Research in the Teaching of English, 1979, 13, 207-216.

Stotsky, S.L. Sentence-combining as a curricular activity: its effects on written language development and reading comprehension. Research in the Teaching of English, 1975, 9, 30-71.

Sullivan, M.A. Parallel sentence-combining studies in grades nine to eleven. In D. Daiker, A. Kerek & M. Morenberg (Eds.) Sentence combining and the teaching of writing. Akron, Ohio: L & S Books, 1979.

Wijnstra, J.M. Syntaktische kompleksiteit in schriftelijk tallgebruik van zesdeklassers. Nederlands Tijdschrift voor de Psychologie, 1972, 27, 29-55.

TABLE 1

Correlations between Quality and Two Measures of Syntactic Complexity in Three Studies

STUDY	LEVEL	WORDS/ T-UNIT	WORDS/ CLAUSE
Faigley	College freshmen	.04 NS	.18 *
Nold and Freedman	College freshmen	- .08 NS	--
Stewart and Grobe	5	.30 *	.23 *
Stewart and Grobe	8	.19 NS	.20 NS
Stewart and Grobe	11	- .06 NS	.12 NS

NS - not significant

* - significant at or beyond the .05 level.